NEURAL INFORMATION PROCESSING SYSTEMS

A Unified Solution for **Privacy and Communication Efficiency** in Vertical Federated Learning

Ganyu Wang, Bin Gu*, Qingsong Zhang,
Xiang Li, Boyu Wang, Charles X. Ling*

Western University, Jilin University, MBZUAI, Xidian University

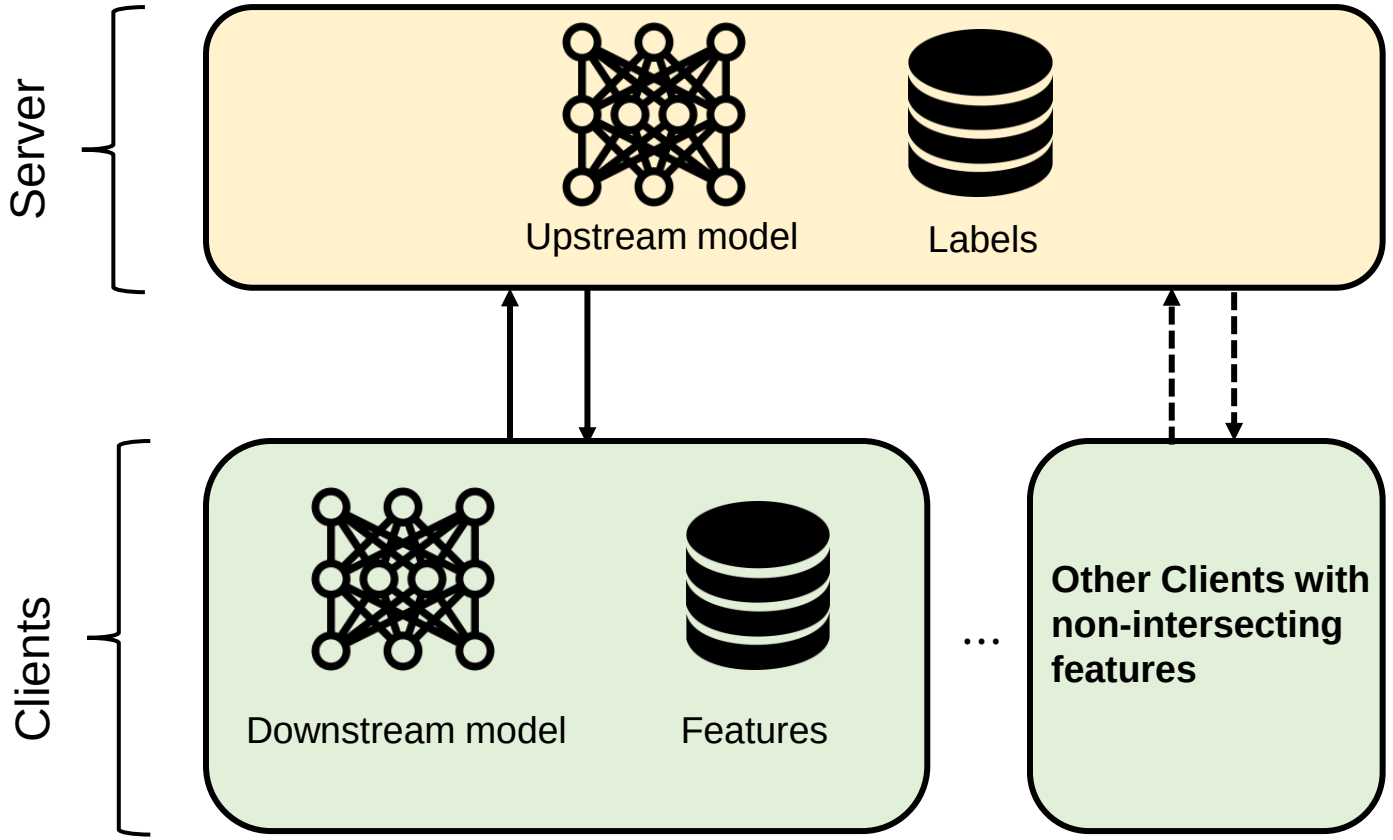
Western

吉林大学

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

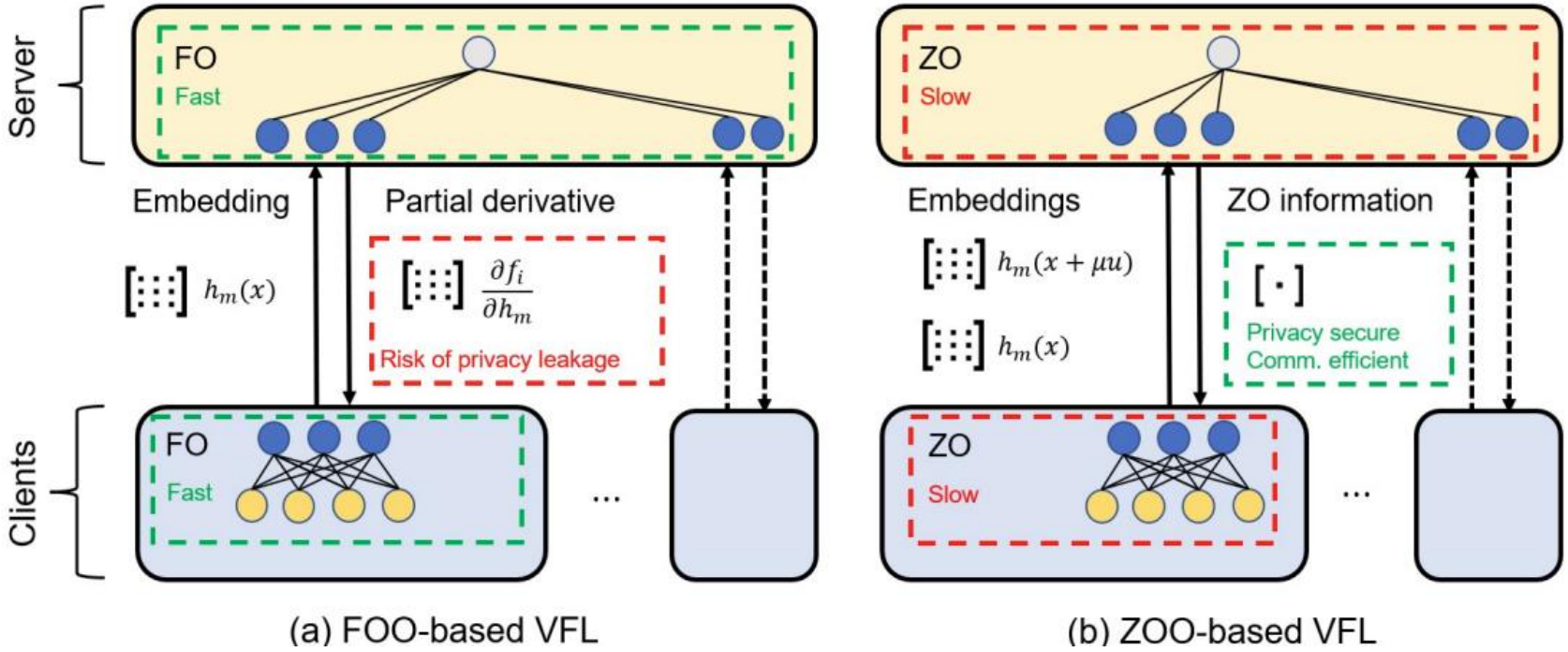
西安電子科技大學
XIDIAN UNIVERSITY

Background: **Vertical** Federated Learning (VFL)



The participants in Vertical Federated Learning (VFL) comprise major organizations, including governments, banks, and companies

Motivation: From First-Order Optimization to **Zeroth-Order Optimization** in VFL



(a) FOO-based VFL

(b) ZOO-based VFL

- Remaining problems for ZOO-based VFL.
  - **Slow convergence** rate limited by the global model size.
  - **Privacy protection** of ZOO has not been theoretically explained.
- We solve the above problems.

Chen, Tianyi, et al. "VAFL: a method of vertical asynchronous federated learning." *arXiv preprint arXiv:2007.06081* (2020).

Zhang, Qingsong, et al. "Desirable companion for vertical federated learning: New Zeroth-order gradient based algorithm." Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021.

Intuition of our Framework.

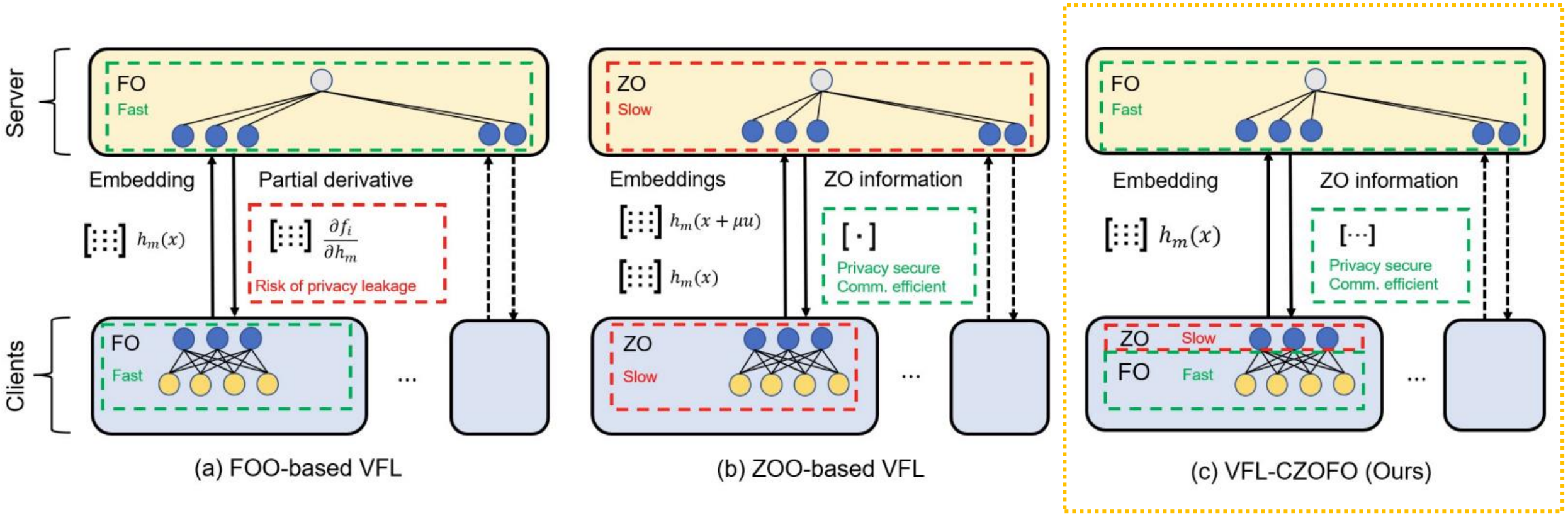


Figure 1: Comparing the Asyn-VFL Frameworks

Problem Definition

- VFL framework can be formalized as a finite-sum problem in composite form:

$$f(w_0, \mathbf{w}; X, Y) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{[\mathcal{L}(w_0, h_{1,i}, \ldots, h_{M,i}; y_i)]}_{f_i(w_0, h_{1,i}, \ldots, h_{M,i})}$$

$$\text{with } h_{m,i} = h_m(w_m; x_{m,i}) \quad \forall m \in [M]$$

Main Solution: Cascaded Zeroth Order Optimization Precisely on the Output Layer of the Clients.

- Estimate the partial derivative of f w. r. t. the **output layer of the clients** with Avg-RandGradEst:

$$\hat{\nabla}_{h_{m,i}} f_i\left(w_0, h_{1,i}, \cdots, h_{M,i}\right) = \frac{\phi(d_{h_m})}{q\mu_m} \sum_{j=1}^{q} \underbrace{[f_i(h_{m,i} + \mu_m u_{m,i}^j) - f_i\left(h_{m,i}\right)]}_{\delta_{m,i}^j} u_{m,i}^j$$

- The estimated partial derivative of f w. r. t. the **clients' model** can be derived via chain rule:

$$\hat{\nabla}_{w_m} f_i\left(\cdot\right) = \hat{\nabla}_{h_m} f_i\left(h_{m,i}\right) \cdot \nabla_{w_m} h_{m,i}$$

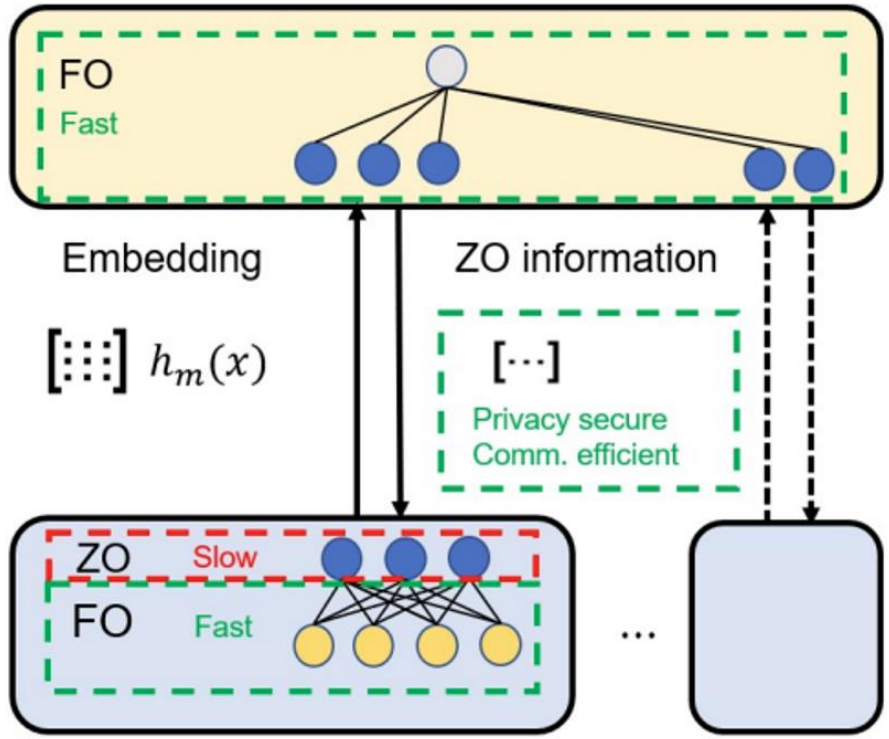
- With this framework, we precisely apply ZOO on the most needed part of VFL.

Algorithm

Algorithm 1 VFL-CZOFO

Input: Learning rate η_m, smoothing parameter μ_m, compressor $\mathcal{C}_m$ for $m \in \{0, 1, ...M\}$.
Output: Parameter w_m for $m \in \{0, 1, ...M\}$.

Initialize model parameter w_m for all participants $m \in \{0, 1, ...M\}$
while not convergent **do**
  when a **client** m is activated, **do**:
    Randomly select a sample $x_{m,i}$
    Compute and send $\mathcal{C}_m(h_m(w_m; x_{m,i}))$ to server
    Receive $\mathcal{C}_0(\Delta_{m,i})$ from the server (in a listen manner)
    Compute $\hat{\nabla}_{h_{m,i}} \hat{f}_i(w_0, \Phi_i)$ via Avg-RandGradEst (Eq. 2)
    Compute $\nabla_{w_m} h_m(w_m; x_{m,i})$ via backpropagation.
    Compute $v_m = \hat{\nabla}_{h_{m,i}} \hat{f}_i(w_0, \Phi_i) \cdot \nabla_{w_m} h_{m,i}$
    Update $w_m \leftarrow w_m - \eta_m v_m$
  when server receives from client m, **do**:
    Compute $\delta^l_{m,i}$ and group into $\Delta_{m,i}$
    Send $\mathcal{C}_0(\Delta_{m,i})$ to the client m
    Compute $v_0 = \nabla_{w_0} f_i(w_0, \Phi_i)$, (FOO)
    Update $w_0 \leftarrow w_0 - \eta_0 v_0$
end while



(c) VFL-CZOFO (Ours)

Convergence Analysis

Theorem 5.2. *Under assumption 5.1, to solve the problem 1 with algorithm 1, the following inequality holds.*

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left\|\nabla f\left(w_0^t,\mathbf{w}^t\right)\right\|^2 \leq & \frac{4p_*\mathbb{E}\left(f^0-f^*\right)}{T\eta}+2\eta p_* L_*\left(4d_*\mathbf{G}_*^4+\mu_*^2L_*^2d_*^2\mathbf{G}_*^2+2\sigma_0^2\right)\\ &+7\eta^2p_*\tau\mathbf{G}_*^2\left(4d_*\mathbf{G}_*^2+\mu_*^2L_*^4d_*^2+2L_*^2\Gamma\right)\\ &+\mu_*^2p_*L_*^2d_*^2\mathbf{G}_*^2+\mathcal{E}p_*H_*^2\left(6+16\mathbf{G}_*^2\right)+17\Gamma p_*\mathbf{G}_*^2\end{aligned}$$

Remark 5.3. If we choose $\eta=\frac{1}{\sqrt{T}}$ and $\mu_*=\frac{1}{\sqrt{T}}$. Design the compression to make $\mathcal{E}=\mathcal{O}\left(\frac{1}{\sqrt{T}}\right)$ and $\Gamma=\mathcal{O}\left(\frac{1}{\sqrt{T}}\right)$ we can derive:

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left\|\nabla f\left(w_0^t,\mathbf{w}^t\right)\right\|^2=\mathcal{O}\left(\frac{d_h}{\sqrt{T}}\right)$$

Remark 5.4. This theorem proves that we significantly relieve the slow convergence problem of ZOO-based VFL [41] from $\mathcal{O}(\frac{d}{\sqrt{T}})$ to $\mathcal{O}(\frac{d_h}{\sqrt{T}})$.

Security Analysis

- We propose a theoretical explanation of the **security of ZOO**. That is ZOO provides **implicit** Differential Privacy (DP) guarantee for sharing the ZO gradient to the client. (Note that the guarantee is intrinsic provided by ZOO itself, instead of adding noise in the mainstream DP mechanism.)

Theorem 4.1. Differential Privacy Guarantee of Sharing the ZO Gradient: Under the "honest-but-colluded" threat model where the attacker can access all information of all clients through the entire training process, including the client's dataset, model, and the ZO gradient received from the server. Under algorithm 1, with q being sufficiently large, ζ being the maximum l_2-norm of the partial gradient w.r.t. any client's output through the entire training. If the following condition holds:

$$\sigma_{m_t,s} = \sqrt{\frac{1}{qd_hT}\sum_{t=0}^{T-1}\mathrm{tr}(\Psi_{m_t}^t)} > \frac{2\zeta\sqrt{2\ln(1.25/\delta)}}{N\cdot\epsilon} \tag{3}$$

we can derive that sharing the stochastic gradient estimation from the server ensures $(\bar{\epsilon}, \bar{\delta})$-DP for all $\epsilon, \delta, \delta' > 0$, where $\bar{\epsilon} = \sqrt{2T\ln(1/\delta')}\epsilon + T\epsilon(e^{\epsilon}-1)$, $\bar{\delta} = T\delta + \delta'$.

Experiment Highlights: Reducing Communication Cost.

Table 3: Evaluating the Test Accuracy and the Comm. Cost

	Privacy	Test Accuracy	Cost(60%)	Cost (80%)	Cost(90%)	Cost (95%)	Cost (total)
Split learning [34]	✗	98.03 ± 0.07	175 MB	234 MB	937 MB	1758 MB	5859 MB
Compressed-VFL [5]	✗	98.17 ± 0.13	109 MB	330 MB	440 MB	1099 MB	3663 MB
VAFL [6]	✗	97.36 ± 0.14	123 MB	184 MB	307 MB	1106 MB	6144 MB
VAFL[6]+DP[1]	✓	95.81 ± 0.20	123 MB	184 MB	799 MB	1413 MB	6144 MB
Syn-ZOO-VFL	✓	84.51 ± 0.72	1406 MB	2520 MB	-	-	-
ZOO-VFL [41]	✓	85.62 ± 0.45	860 MB	1300 MB	-	-	-
Ours	✓	95.00 ± 0.21	**16** MB	**24** MB	**39** MB	**205** MB	790 MB

Experiment Highlights: Intrinsic DP guarantee of ZOO

- DP guarantee for the ZOO (Avg-RandGradEst) with different sampling times.

Table 2: Sampling Times of Avg-RandGradEst and the Corresponding Differential Privacy Guarantee

q	σ^2	ζ	δ	ϵ	δ'	$\bar{\epsilon}$	$\bar{\delta}$
10	1.61	20.1	10^{-6}	2.8×10^{-3}	10^{-6}	85.9	9.4×10^{-2}
100	0.15	7.3	10^{-6}	3.3×10^{-3}	10^{-6}	93.6	9.4×10^{-2}

Conclusion

- 1 We propose a **cascaded hybrid optimization VFL framework** that improves communication efficiency and privacy simultaneously.
- 2 Theoretical analysis shows:
  - 1) the intrinsic (ϵ, δ)-differential privacy guarantee provided by ZOO within our framework,
  - 2) the convergence of our framework and its superiority compared with the ZOO-based VFL,
- 3 Experiment on differential privacy and communication efficiency.

Thank you very much!